

Das Informationschaos in der Netz-Welt ordnen

Von Ludwig Siegele • Der Amerikaner **Brewster Kahle** hat eine neue Internet-Navigationshilfe erfunden. Ende des Such-Frusts?

Der Metallkasten ist nicht grösser als ein Kleiderschrank. Aber er hat es in sich: Fast das ganze Internet ist darin gespeichert – Millionen von Fotos, Grafiken und Texten. Sie füllen mehrere hundert Datenkassetten, die ein Roboterarm immer wieder aus einem Regal herausnimmt, in ein Lesegerät steckt und dann zurücklegt.

«Zwanzig Terabyte, also zwanzig Millionen Megabyte, passen da rein – etwa soviel Informationen wie in allen Büchern der Library of Congress stecken, der grössten Bibliothek der Welt», erklärt Brewster Kahle, der Herr der gesammelten Netz-Werke. «Ich kann es immer noch nicht fassen, dass dies technisch möglich ist. Das ist phantastisch!»

Wenn Kahle auf sein jüngstes Projekt angesprochen wird, dann ist er kaum noch zu bremsen: 1996 begann er, das Internet komplett zu archivieren. Jetzt will er auch Geld damit verdienen. Alexa heisst sein Unternehmen (www.alexaproject.org) – eine Navigationshilfe neuen Typs, die endlich das Informationschaos in der Netzwelt etwas ordnen könnte.

Kahle ist auch keiner der vielen auf ihren Börsengang fixierten Jungunternehmer in und um San Francisco, sondern ein Mann mit Mission und dafür in der Netzgemeinde wohlbekannt: «Unsere Technologie soll helfen, dass Menschen besser und schneller kommunizieren können. Das macht Gemeinschaften intelligenter.»

Von solch hochgesteckten Zielen liess sich der heute 37jährige Kahle schon während seines Informatikstudiums am Massachusetts Institute of Technology (MIT) leiten: Er beschäftigte sich vor allem mit Techniken, mit denen sich riesige Datenmengen in Sekunden durchforsten lassen – et-

«Unsere Technologie wird den Menschen helfen, nicht nur besser, sondern auch schneller miteinander zu kommunizieren. Das macht Gemeinschaften intelligenter.»

wa um Informationen zu finden, aber auch um Zusammenhänge zwischen ihnen herzustellen.

In den achtziger Jahren baute Kahle bei der Firma Thinking Machines Supercomputer, in denen Tausende von Prozessoren parallel rechnen. Und Anfang der neunziger Jahre entwickelte er WAIS, eine der besten Suchtechnologien für das Netz und Namensgeber für sein erstes Unternehmen, das er 1995 an den Online-Dienst AOL verkaufte.

Für Kenner der Online-Szene war es denn auch keine Überraschung, dass gerade Kahle einer der ersten war, die Antwort auf eine drängende Frage suchten: Wie lässt sich das Netz archivieren? «Für digitale Informationen gibt es bisher kein Äquivalent einer Bibliothek», schrieb er schon im März 1996 im Wissenschaftsmagazin «Scientific American». Wenn sich daran nichts ändere, drohe der Menschheit ein unwiederbringlicher Verlust.

Um die Gefahr abzuwenden, starteten Kahle und einige Freunde im gleichen Jahr das Internet-Archiv (www.archive.org). Ihr Ziel: das Netz so komplett wie möglich abzuspeichern, also nicht nur sämtliche öffentlich zugänglichen Webseiten, sondern auch die elektronischen Diskussionsforen, Usenet Groups genannt.

Aber Kahle ging es von Anfang an nicht nur darum, die Datenmassen sicher abzulegen, sondern sie auch nutzbar zu machen. Deswegen gründete er vor wenigen Monaten Alexa, be-

nannt nach der alexandrinischen Bibliothek, der bedeutendsten Sammlung von Dokumenten der Antike, die bei der Belagerung der ägyptischen Hauptstadt durch Cäsar abbrannte.

«Wir treiben die Maschinen hier zum Äussersten. Dafür sind sie eigentlich gar nicht gemacht», sagt Kahle und zeigt auf den Speicherschrank in Alexas Hauptquartier mitten im Presidio Park in San Francisco, gleich neben der Golden-Gate-Brücke. An dem Gerät macht sich gerade ein Techniker zu schaffen. Der Roboterarm hat mal wieder den Geist aufgegeben.

Festplatten, wie sie in jedem Personalcomputer stecken, wären sicher weniger anfällig. Aber sie wären auch weit teurer – kein unwichtiger Gesichtspunkt, bei den Datenmengen, die bei Alexa anfallen: Rund 200 Dollar kostet es, ein Gigabyte (tausend Megabyte) auf dem Medium abzuspeichern. Die Datenkassetten mit dem Roboterarm schlagen dagegen nur mit zwanzig Dollar pro Gigabyte zu Buche.

In einer anderen Ecke des Unternehmens drehen sich freilich einige Dutzend Festplatten, auf denen sich die Musikkonserven einer durchschnittlich grossen Radiostation bequem unterbringen lassen würden. Hier, im Gehirn des Unternehmens, wird kräftig «datamining» betrieben: Das Gold wird aus Alexas riesigem Datenbergwerk herausdestilliert.

Die fetten Festplatten und die dazugehörigen schnellen Rechner liefern Kahle Zahlen, die sonst keiner hat: Die durchschnittliche Lebensdauer einer Seite im Web ist vierundvierzig Tage. Ihre Zahl verdoppelt sich alle sechs Monate. Derzeit zählen sie zusammen zwei Terabyte – genug, um die Festplatten von tausend durchschnittlich leistungsstarken Personalcomputern zu füllen.

Aber die Maschinen schaffen vor allem die Grundlage für das wichtigste Angebot von Alexa: Netznutzern mit intelligenten Navigationsvorschlägen zu helfen, jene Webdienste im elektronischen Weltgespinnst zu finden, die sie auch wirklich interessieren.

«Die Anlage hier ordnet Webdienste in Gruppen ein, in denen beispielsweise bestimmte Wörter besonders häufig vorkommen», erklärt Kahle. «Angenommen, Sie besuchen gerade einen Internet-Service, der Rucksackreisen verkauft. Dann schlägt Ihnen Alexa andere Offerten vor, die sich ebenfalls mit dem Thema beschäftigen.»

Auf die Dauer sollen die Nutzer des Dienstes freilich vor allem selbst bestimmen, wohin sie geleitet werden. Denn die Rechner von Alexa merken sich, welche virtuellen Wege Websurfer gehen. Kahle: «Wenn Sie auf einem Trampelpfad durch den Wald gehen, dann profitieren Sie von

den Exkursionen Ihrer Vorgänger. Genau dieses Prinzip wenden wir auf das Netz an.»

Wer sich von Alexa durch das Netz führen lassen will, muss sich ein kleines, kostenloses Programm vom Webdienst des Unternehmens herunterladen. Es erweitert die Browser-Software von Netscape oder Microsoft um eine Menüleiste am unteren Rand des Bildschirms. Darauf klicken die Nutzer, wenn sie die Surf-Empfehlungen sehen wollen.

Dort finden sich freilich noch andere interessante Buttons – etwa jener, der ein Fenster mit Informationen über den gerade besuchten Dienst öffnet: Wer betreibt den Service? Wie beliebt ist er? Wie vielen anderen Webangeboten ist er einen Link wert, einen elektronischen Querverweis? Beachtet er die Regeln der freiwilligen Datenschutz-Initiative TrustE (www.truste.org)?

«Das eine grosse Problem des Internets ist, Informationen zu finden», begründet Kahle die Fülle von Angaben über einen Dienst. «Die andere, mindestens ebenso grosse Schwierigkeit besteht darin, dass man nie genau weiss, wo man landet. Wie wollen Sie ohne solche Metadaten herausfinden, ob Sie etwa einem Online-Reisebüro trauen können oder nicht?»

«Wenn Sie einem Trampelpfad im Wald folgen, dann profitieren Sie von den Exkursionen Ihrer Vorgänger. Genau dieses Prinzip wenden wir auf das Netz an.»

Auch eine dritte Schwäche des Web versucht Alexa zu beheben: Seiten zu ersetzen, die längst gelöscht sind. Wenn der Browser mal wieder «404 document not found» meldet, dann schaut der Dienst nach, ob das angeforderte Dokument im Archiv steckt. Es kann dann per Mausklick auf die Alexa-Menüleiste aufgerufen werden.

Damit nicht genug. Sozusagen als Überraschungsbomben hat Kahle seine Software mit der Encyclopaedia Britannica verbunden: Nutzer können auf einer Webseite einen Begriff markieren und sich ihn dann von der Online-Ausgabe der berühmten Enzyklopädie erklären lassen.

Für sein umfassendes Angebot erntet Kahle allerdings nicht nur Lob. Kritische Geister wenden ein, dass sein Dienst Urheberrechte verletzt und vor allem wenig Rücksicht auf den Datenschutz nimmt: Mancher Netzbürger wolle nicht mehr wahrhaben, was er einmal auf seiner persönlichen Homepage von sich gegeben hat – doch in Alexas Datenbanken seien die Daten noch zu haben.

Kahle ficht diese Kritik nicht an. «Unsere Regel ist klar», erwidert er: «Wer sich in seinem Copyright verletzt fühlt oder Informationen aus der Welt schaffen will, kann uns anweisen, die entsprechenden Dateien zu löschen. Aber bisher haben das nur sehr wenige getan. Im Gegenteil – sie wollen Teil der digitalen Geschichte sein.»

Stärker treffen ihn schon Bemerkungen, seine Computer machten oft unsinnige Surf-Empfehlungen, etwa wenn sie einem Leser der «Weltwoche» im Internet empfehlen, es doch einmal bei einem Webdienst über Schweizer Käse zu versuchen. «Je mehr Menschen das Angebot nutzen, desto besser wird es», meint Kahle dazu.

Richtig ernst nimmt der Alexa-Chef Zweifel von Experten, dass er mit seinem Angebot Geld verdienen kann. Wie bei den meisten anderen Webdiensten auch soll Werbung für die nötigen Einnahmen sorgen: Neben den Surf-Empfehlungen steht eine kleine Anzeige – speziell zugeschnitten auf das, was Netznutzer gerade betrachten.

«Als einzige Geldquelle reicht das auf die Dauer nicht», meint Jerry Michalski, Chefredaktor des renommierten Branchenbriefes «Release 1.0». Bei Geoffrey Bock, Consultant beim Marktforscher Patricia Seybold Group, klingt das ähnlich: «An seinem Geschäftsmodell muss Alexa noch etwas arbeiten.»

Kahle sucht auch schon nach weiteren Einnahmemöglichkeiten. Er will seine Technologie Grossunternehmen andienen, deren interne Netze mittlerweile fast genauso unübersichtlich geworden sind wie das Internet. Und er überlegt, zahlungskräftigen Webdiensten Studien darüber anzubieten, wie ihre Angebote wahrgenommen werden.

Die finanziellen Fragen halten den Internet-Archivar aber nicht davon ab, über die technische Zukunft seines Dienstes nachzudenken. Sein Ziel: eine Art virtuelle Zeitmaschine. Nutzer sollen bei ihm abfragen können, wie ein Webangebot zu einem bestimmten Zeitpunkt ausgesehen hat.

So will er seine Technologie immer weiter verbessern. Einmal, so hofft Kahle, wird sie den Menschen sogar helfen, die schweren Fragen des Lebens zu beantworten: Welches Buch soll ich lesen? Soll ich auf die Universität gehen? Wie soll ich meine Kinder erziehen? «Ich will», träumt er, «ein richtiges Orakel von Delphi schaffen.»

Der Herr der gesammelten Netzwerke:
Brewster Kahle



every day as if for work, stretched to pay their Catholic-school tuition and enlisted teachers to help keep his firing a secret. "Kids need to focus on going to school and getting an education," he says.

Today Mr. Bryant has few regrets. He feared losing custody if he didn't meet high parenting standards, and "keeping my kids was more important than anything in the world."

VIVIAN MARTYSZENKO'S loss of both her job and a federal lawsuit against her employer began a period of near-chaos for her family.

As reported, Ms. Martyszenko was replaced on her job as a Safeway cashier in Ogallala, Neb., after missing work for weeks while her family was ensnarled in a 1995 police investigation. Anonymous charges that her children had been sexually molested by someone outside the family were eventually found groundless and dropped. But it was too late for Ms. Martyszenko, who had feared, based on an investigator's comment, that officials would take her children away from her if she left them with a sitter to go to work.

Safeway says it did everything it could, giving her two weeks' vacation, offering to schedule her shifts around appointments with an investigating psychiatrist and even offering her her job back in an attempt to set-

His federal sex-discrimination suit against the Maryland State Police is set for trial in January. As reported, his fight for the 30 days' paid childbirth leave given to "primary caregivers" under Maryland law has drawn national attention. Police officials vigorously deny the discrimination charge, saying leave has always been fairly granted to both sexes.

Regardless of the outcome, Mr. Knussman already sees benefits from turning a spotlight on the problems of working fathers. After the 1996 birth of his second child, he received the 30 days' paid leave he requested. Maryland has changed an ambiguous employee paid-leave law so that it clearly applies to all caregivers.

His effort to avert an adverse transfer proposed by his bosses, which would have tripled his commute time, has succeeded so far. Though the court threw out his claim that the transfer was in retaliation for his lawsuit, his bosses haven't pursued the proposed transfer.

Nevertheless, when asked if he would begin his battle again, Mr. Knussman pauses for a full 10 seconds and sighs. "The jury's still out on that one," he says. The pending lawsuit "has had a pretty significant impact on me and my family," he adds, causing stress for his wife and stirring fears that he'll lose his job. The worry "is kind of like a ringing in your ear: It's always there."

By CHRISTINA BINKLEY
Staff Reporter of THE WALL STREET JOURNAL

LAS VEGAS—Steve Wynn is famous throughout the casino industry for micro-managing Mirage Resorts Inc., right down to the flowers at the front desk. But even by Chairman Wynn's own standards, the attention he is lavishing on the final touches of his new Bellagio casino is exceptional.

Costing \$1.9 billion, Bellagio is the most expensive casino ever, complete with mini-version of Italy's Lake Como. For tomorrow's opening, Mr. Wynn chose the orange-threaded Japanese paper for the gala invitations and has been tasting menu



Steve Wynn

What Did You Do Online, Daddy? Preserving the Web for Posterity

By KARA SWISHER

Staff Reporter of THE WALL STREET JOURNAL

Is the World Wide Web already a museum piece?

It seemed so yesterday, when a chunk of the Internet, circa 1997, was given to the Library of Congress by an unusual San Francisco company that for two years has tried to capture bits and bytes before they dissipate into the digital ether.

Forty-four digital tapes encased in an interactive sculpture will take up residence with millions of books, films, sound recordings and other material the library has collected for decades. Titled "World Wide Web 1997: 2 Terabytes in 63 inches," the material amounts to about one-tenth of the data already stuffed in the library's stacks, in a package about as big as a large file cabinet.

The sculpture has four red computer monitors that flash 10,000 Web pages—two every second—from the 500,000 sites collected by Alexa Internet since early 1997.

It's all the brainchild of an enthusiastic techie named Brewster Kahle, who founded Alexa after selling his first Internet company to America Online Inc. in 1995 for \$15 million in stock. With the windfall, he helped start Internet Archive, a Seattle nonprofit, to preserve copies of the Web for research purposes and posterity. After Alexa "crawls" the public Web using robot

technology every few weeks, it donates a copy to the archive. It's now 12 terabytes in size, equivalent to 12 million books.

"The Web is being born and destroyed every day, and we can't miss the chance to capture what it was like as it is on its way to being something else," says Mr. Kahle, who named his company after the Library of Alexandria in ancient Egypt and likes to think of himself as the Internet's Demetrius, the first librarian there. But with the Web growing by an estimated 1.5 million pages a day, coming up with a way for people to visualize its scope was tricky. "We did not want to make it too overwhelming, and we also did not want to make it look too dinky," he says.

Library of Congress associate librarian Winston Tabb thinks Mr. Kahle got it just right. "We felt we needed to get in on the ground floor and capture this material now, even though we don't know what we are going to do with it," says Mr. Tabb. "It's a shame that so much has already been lost from the Web's earliest years."

Mr. Kahle attributes that to the medium's fast-changing nature and says that some Web site creators were reluctant to let him even take snapshots. "What they don't realize is that they are history if they are not part of history," he says. "You push a button and erase it and it's gone."

items for the evening's repast. The year-old chairman has also been fussy over such details as the fringes on curtains and the lighting on decorative sculpture one of Bellagio's restaurants.

"He's the most motivating man to work for—also terrifying," says Mark Full, president and chief executive of Wet sign, a Universal City, Calif., company that designed the \$30 million dancing fountains beckoning visitors into Bellagio.

As Mr. Wynn fusses, an entire industry is holding its breath. Bellagio is set to be upon the scene amid a round of extravagant parties, some lasting three days. Some million is being spent to market and celebrate Bellagio's first weeks, according to Lee, Mirage Resorts' chief financial officer. That includes 60-second TV ads playing nationally and as far away as Hong Kong.

Mr. Wynn believes an infusion of hoopla may be just the right tonic for enervated gambling capital. In his estimation, Las Vegas has tapped out traditional casino demand world-wide and must appeal to the sophisticated, well-traveled consumers who have snubbed the pla-